

# Arabidopsis

## A Model Plant Genome

**A**n ambitious international venture to decode nearly all of the genetic material in a flowering plant called *Arabidopsis thaliana*—a member of the mustard family—is speeding ahead of schedule. And the project is already hastening the discovery of important genes in crop plants.

*A. thaliana*, also known as mouse-ear cress or thale cress, has become a workhorse of plant biotechnology because it has only a small amount of genetic material. A small genome helps simplify the task of determining the exact sequence, or order, of the four chemical subunits that make up genetic material or DNA. Every plant gene is composed of a unique sequence of at least a thousand of these subunits, or bases.

Scientists sleuthing *Arabidopsis* originally estimated that they would finish their precedent-setting task by 2004, says molecular biologist Athanasios Theologis. He directs the *Arabidopsis* sequencing project at the ARS/University of California at Berkeley Plant Gene Expression Center in Albany, California. Now, researchers expect to finish by 2000, giving the world what has been described as “the first essentially complete catalog of all the genes involved in the life cycle of the typical plant, from seed to flower to fruit.”

Theologis and colleagues at Stanford University in Stanford, California, and the University of Pennsylvania in Philadelphia use an approach called high-throughput DNA sequencing. It is among the newest, cheapest, and fastest ways to find out the sequence of base units on the long strands of tightly coiled DNA in the chromosomes of *Arabidopsis*.

Once the sequence of bases is known, it is promptly posted on GenBank, an up-to-the-minute Internet compilation that also displays mouse, human, and other sequences. Scientists can use GenBank to look for sequences other organisms have in common. Because similarities in

Theologis and co-researchers have already discovered about 500 *Arabidopsis* genes. Currently, the team sequences about one gene a day. High-throughput DNA sequencing offers that payoff because it focuses on finding sequences first and leaves the question of a gene’s function for later.

In contrast, the conventional approach begins by investigating a function or trait first, then finds the gene responsible for the trait, and finally decodes the gene’s sequence. That one-by-one approach, says Theologis, costs millions, while finding a gene with high-throughput DNA sequencing averages only a modest \$2,000 per gene. Theologis expects the price to drop even lower.

The National Science Foundation, U.S. Department of Energy, and ARS are funding the work at Albany. Foundation and Energy Department funds are, in addition, going to Stanford, Philadelphia, and four other U.S. labs. Scientists in Japan and Europe are also part of the quest to unlock all of the sequences in the *Arabidopsis* genome.—By **Marcia Wood, ARS.**

*Athanasios Theologis is at the USDA-ARS/University of California at Berkeley Plant Gene Expression Center, 800 Buchanan St., Albany, CA 94710; phone (510) 559-5911, fax (510)*

*559-5678, e-mail theo@nature.berkeley.edu.*

*Visit the Plant Gene Expression Center’s Genome Sequencing Laboratory on the WWW at <http://pgec-genome.pw.usda.gov>.* ♦

KEITH WELLER (K3954-7)



Tissue-cultured *Arabidopsis thaliana*.

sequence often indicate similarities in function, searching for sameness via computer quickens the discovery of a gene’s function.

Once biotechnologists discover the sequence and function of a gene, they may be able to move useful genes into plants that lack them, such as a gene for disease resistance, for example. Or they might be able to rebuild a gene to boost its effectiveness.